CROSS-VARIETAL PATTERNS IN THE ENGLISH GENITIVE ALTERNATION: A PILOT STUDY

ICAME 36, Trier

May 30, 2015

Benedikt Heller
benedikt.heller@kuleuven.be

OVERVIEW

1. Introduction
2. Predictors of the genitive alternation
3. Data
4. Regression model

INTRODUCTION

PROJECT

- “Exploring probabilistic grammar(s) in varieties of English around the world”
- 5-year project (2013–2018), KU Leuven
- Three alternations
  - Particle placement: Dr. Jason Grafmiller
  - Dative alternation: Melanie Röthlisberger
  - Genitive alternation: Benedikt Heller
- Supervisors of PhD project
  - Prof. Dr. Benedikt Szmrecsanyi
  - Prof. Dr. Joybrato Mukherjee
  - Dr. Jason Grafmiller

THE GENITIVE ALTERNATION



(1) [The family]$_{possessor}$'s [spokesperson]$_{possessum}$

(2) The [spokesperson]$_{possessum}$ of [the family]$_{possessor}$

- "Today, genitive variation is arguably the best researched of all syntactic alternations in English" (Rosenbach 2014)

INTERCHANGEABILITY

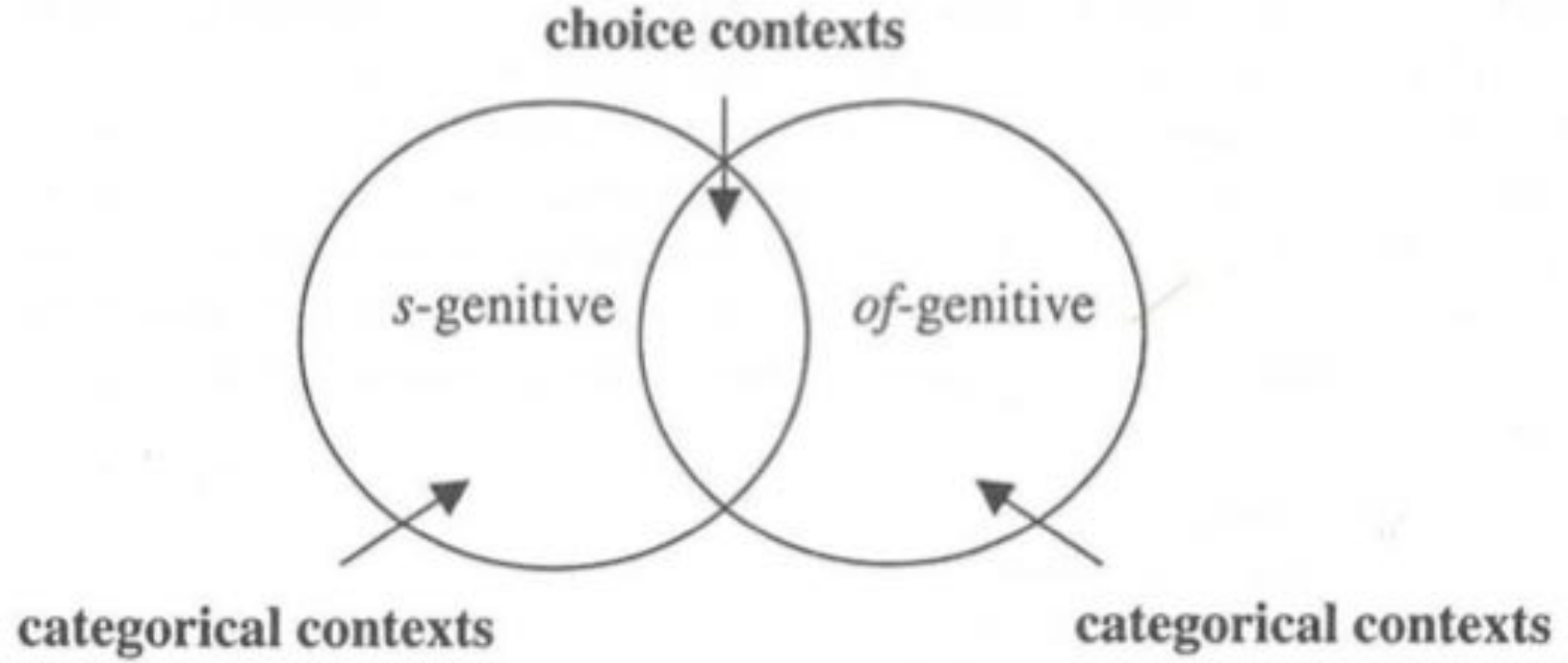


Figure 1. Categorical versus choice contexts

(Rosenbach 2002: 28)

INTERCHANGEABILITY

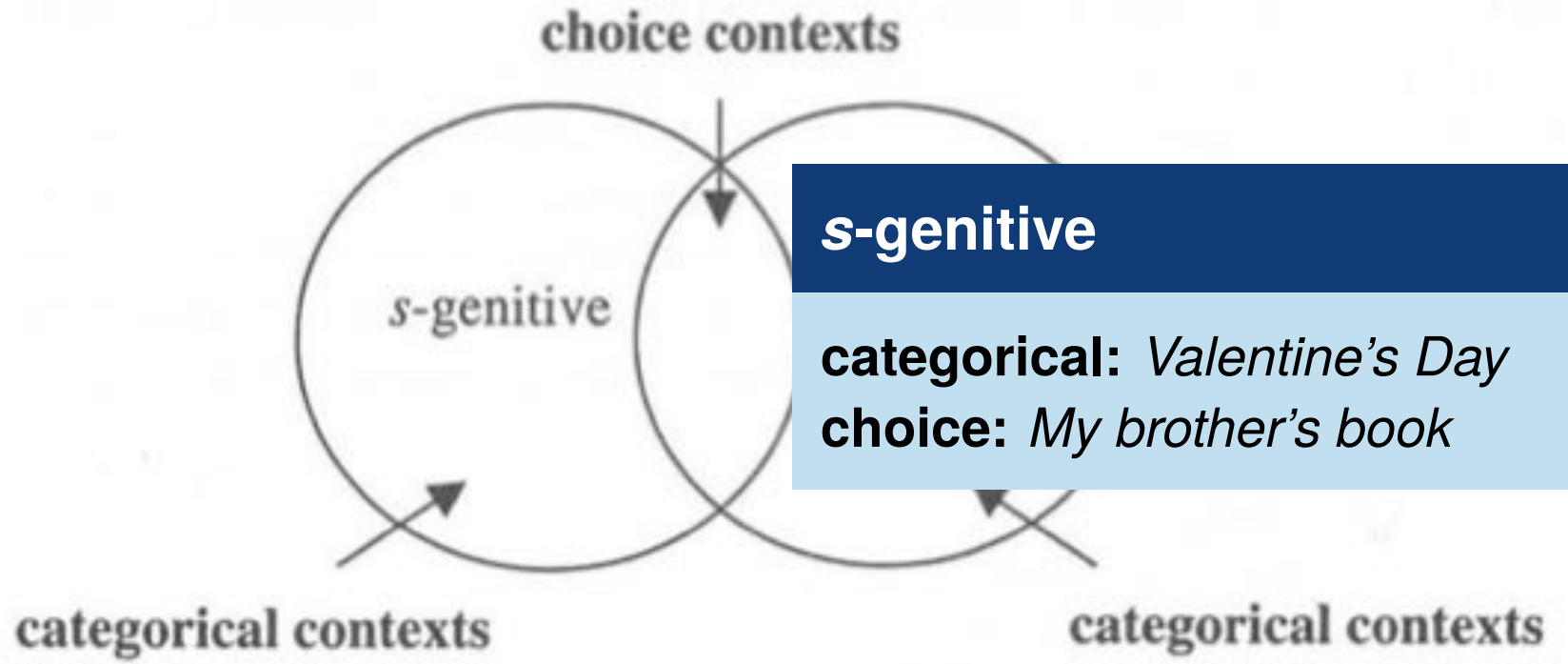


Figure 1. Categorical versus choice contexts

s-genitive

categorical: *Valentine's Day*
choice: *My brother's book*

(Rosenbach 2002: 28)

INTERCHANGEABILITY

_of_-genitive

categorical: *a couple of years*
choice: *the name of my cat*

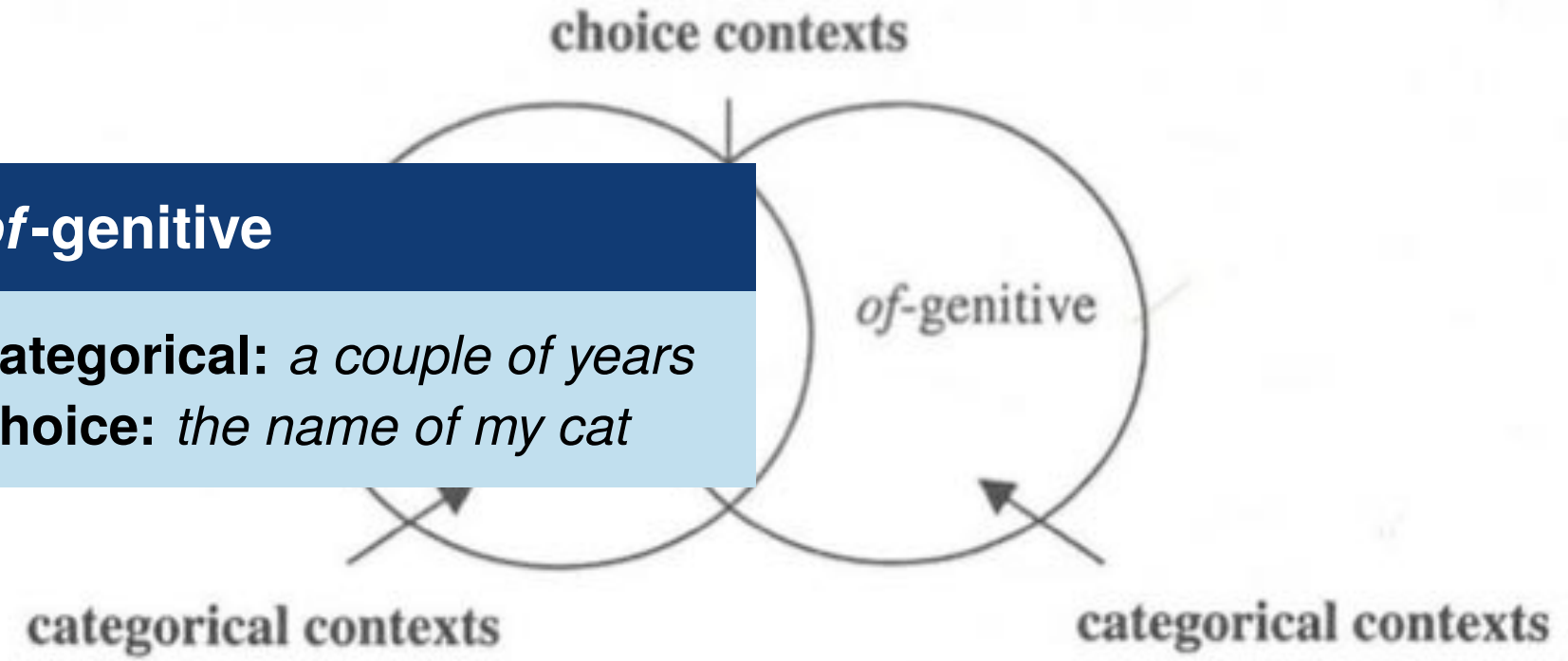
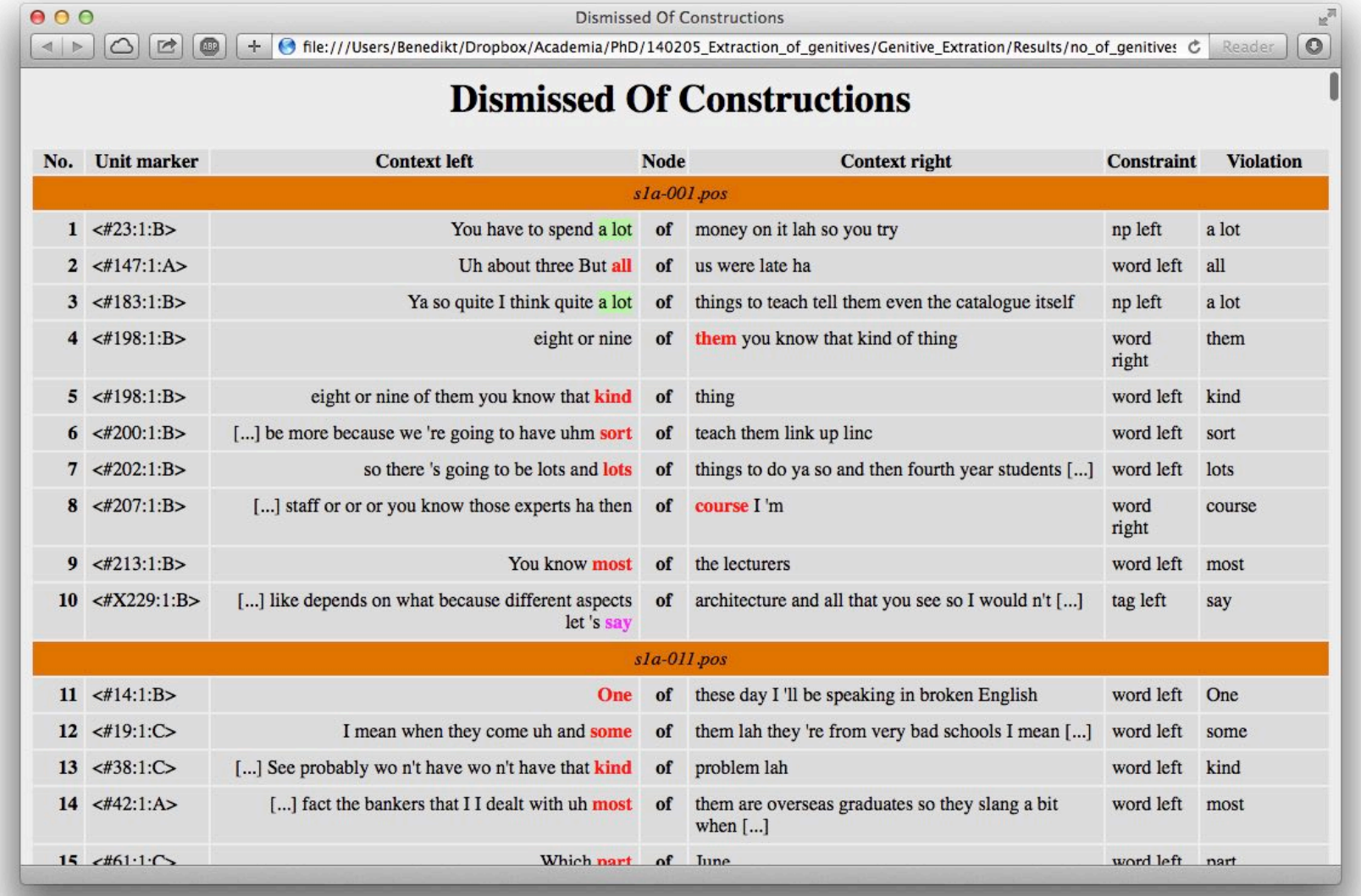


Figure 1. Categorical versus choice contexts

(Rosenbach 2002: 28)

INTERCHANGEABILITY



Dismissed Of Constructions

No.	Unit marker	Context left	Node	Context right	Constraint	Violation
			s1a-001.pos			
1	<#23:1:B>	You have to spend a lot	of	money on it lah so you try	np left	a lot
2	<#147:1:A>	Uh about three But all	of	us were late ha	word left	all
3	<#183:1:B>	Ya so quite I think quite a lot	of	things to teach tell them even the catalogue itself	np left	a lot
4	<#198:1:B>	eight or nine	of	them you know that kind of thing	word right	them
5	<#198:1:B>	eight or nine of them you know that kind	of	thing	word left	kind
6	<#200:1:B>	[...] be more because we 're going to have uhm sort	of	teach them link up linc	word left	sort
7	<#202:1:B>	so there 's going to be lots and lots	of	things to do ya so and then fourth year students [...]	word left	lots
8	<#207:1:B>	[...] staff or or or you know those experts ha then	of	course I 'm	word right	course
9	<#213:1:B>	You know most	of	the lecturers	word left	most
10	<#X229:1:B>	[...] like depends on what because different aspects let 's say	of	architecture and all that you see so I would n't [...]	tag left	say
			s1a-011.pos			
11	<#14:1:B>	One	of	these day I 'll be speaking in broken English	word left	One
12	<#19:1:C>	I mean when they come uh and some	of	them lah they 're from very bad schools I mean [...]	word left	some
13	<#38:1:C>	[...] See probably wo n't have wo n't have that kind	of	problem lah	word left	kind
14	<#42:1:A>	[...] fact the bankers that I I dealt with uh most	of	them are overseas graduates so they slang a bit when [...]	word left	most
15	<#61:1:C>	Which part	of	June	word left	part

RESEARCH QUESTIONS

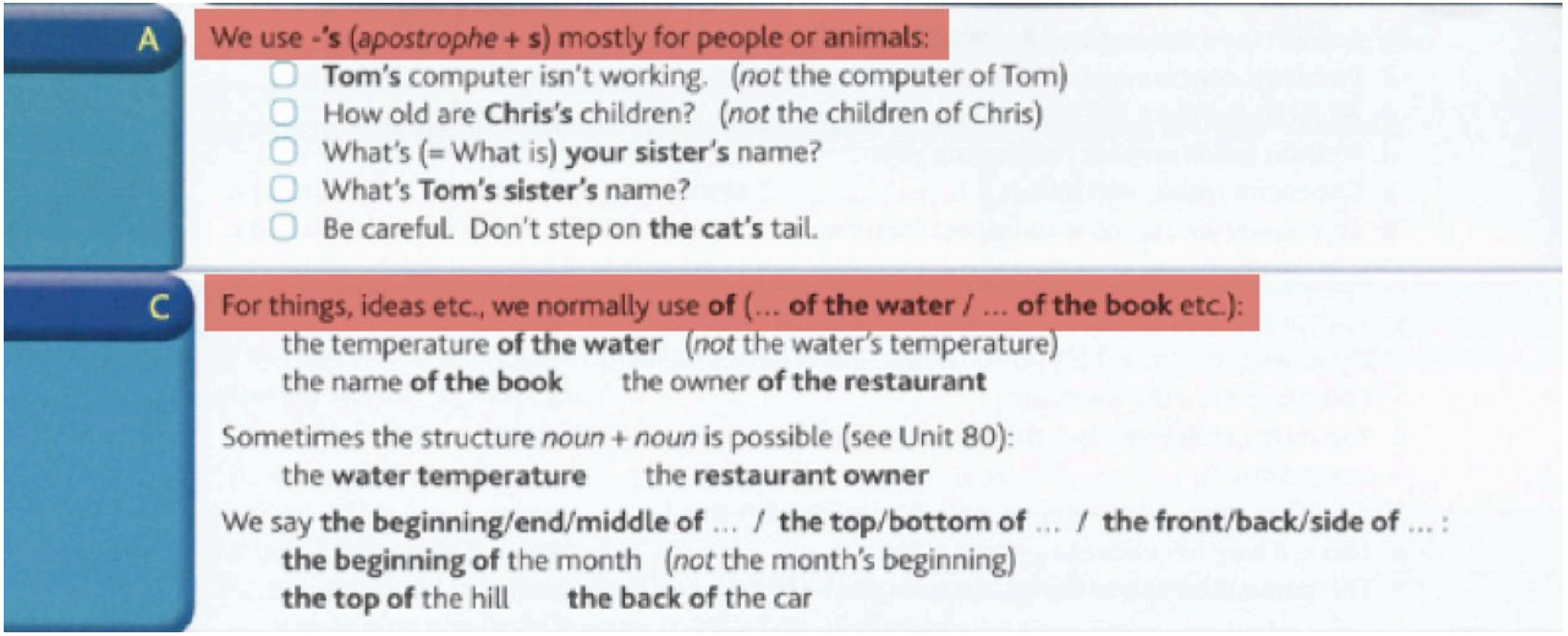
1. Does genitive choice differ across varieties?
2. Does the factor VARIETY interact with well-known predictors?
3. Are cross-varietal patterns congruent with Schneider's (2007) Dynamic Model?

PREDICTORS OF THE GENITIVE ALTERNATION

ANIMACY

- Animacy of the possessor: Most important predictor (Grafmiller 2014; Hinrichs and Szmrecsanyi 2007)
- Prescriptive rule in grammar books (e.g. Murphy 2012)
  - Animate possessor: *s*-genitive
  - Inanimate possessor: *of*-genitive
- Automatic annotation fairly accurate (~90%)

ANIMACY



A

We use **-'s** (*apostrophe* + **s**) mostly for people or animals:

- **Tom's** computer isn't working. (*not* the computer of Tom)
- How old are **Chris's** children? (*not* the children of Chris)
- What's (= What is) **your sister's** name?
- What's **Tom's sister's** name?
- Be careful. Don't step on **the cat's** tail.

C

For things, ideas etc., we normally use **of** (... **of the water** / ... **of the book** etc.):

the temperature **of the water** (*not* the water's temperature)
the name **of the book** the owner **of the restaurant**

Sometimes the structure *noun* + *noun* is possible (see Unit 80):

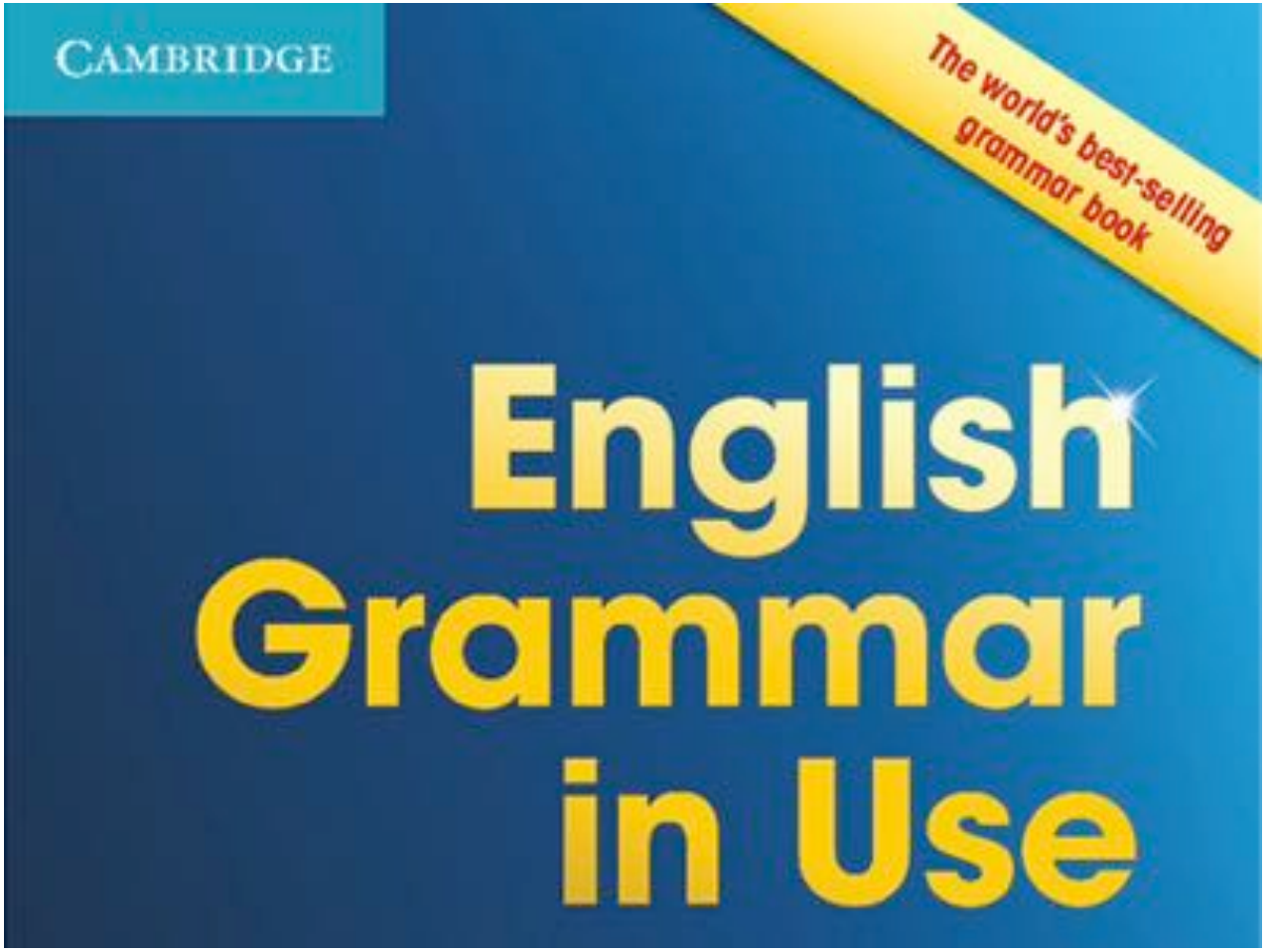
the **water temperature** the **restaurant owner**

We say **the beginning/end/middle of** ... / **the top/bottom of** ... / **the front/back/side of** ... :

the beginning of the month (*not* the month's beginning)
the top of the hill **the back of** the car

(Murphy 2012: 162)

ANIMACY



(Murphy 2012)

FINAL SIBILANCY

- If possessor ends in [s], [z], [ʃ], [tʃ], [ʒ], or [dʒ], e.g.

  (3) The paradox's conclusion <ICE-IND:W2B-021>

  (4) the church's solidarity with women <ICE-NZ:S1B-011>

- Makes *s*-genitive less likely
- Automatically annotated using the CMU Pronunciation Dictionary

```
88426   PARADOX  P EH1 R AH0 D AA2 K S
88427   PARADOXES  P EH1 R AH0 D AA2 K S IH0 Z
88428   PARADOXICAL  P EH2 R AH0 D AA1 K S IH0 K AH0 L
88429   PARADOXICALLY  P EH2 R AH0 D AA1 K S AH0 K L IY0
```

THEMATICITY & GIVENNESS

- High thematicity (i.e. frequency of the possessor in a text) makes *s*-genitive realization more likely
- If a possessor is *given* (i.e. mentioned before) it is also more likely to be realized as *s*-genitive
- Automatic annotation

Beer Glassware

In Belgium, every beer is served in its own signature glass designed to highlight the beer's special flavor. The Belgian Pouring Ritual and Chalice create a unique connection between the beer and customer adding to a fine dining experience. It brings sophistication to the beer-drinking experience and an enhanced appreciation for the serving process that wine drinkers have been enjoying for years. The most well-known Belgian glass is the Stella Artois Chalice, which releases the beer's flavor and multi-textured aromas. Hoegaarden is offered in a hexagonal glass that showcases its natural cloudiness and spicy coriander aroma. The thickness of the glass also helps maintain the desired serving temperature. Leffe uses a high-stemmed glass, enabling customers to enjoy the beer's delicate malt aroma and robust, creamy warmth.

+ Click image to enlarge

SYNTACTIC WEIGHT

- Length of the constituents (possessor, possessum)
- Principle of end weight: "the tendency for long and complex elements to be placed towards the end of a clause." (Biber et al. 1999: 898)
  - Long possessor: *of*-genitive
  - Long possessum: *s*-genitive

(5) the power of the Chinese Government <ICE-HK:S1A-021>

(6) laser's potential medical uses <ICE-IND:W2B-031>

- Measured automatically
- Logged frequency of characters

LEXICAL DENSITY

- Type-token ratio
- Highly sensitive to text length. Therefore, measured for the immediate environment of 100 words (Hinrichs and Szmrecsanyi 2007)
- *S*-genitive realization is more likely in lexically dense environments (e.g. Grafmiller 2014)

OTHER FACTORS

Not significant

- Mode (spoken vs. written)
- Total frequency of head (GloWbE)

Not included in present analysis

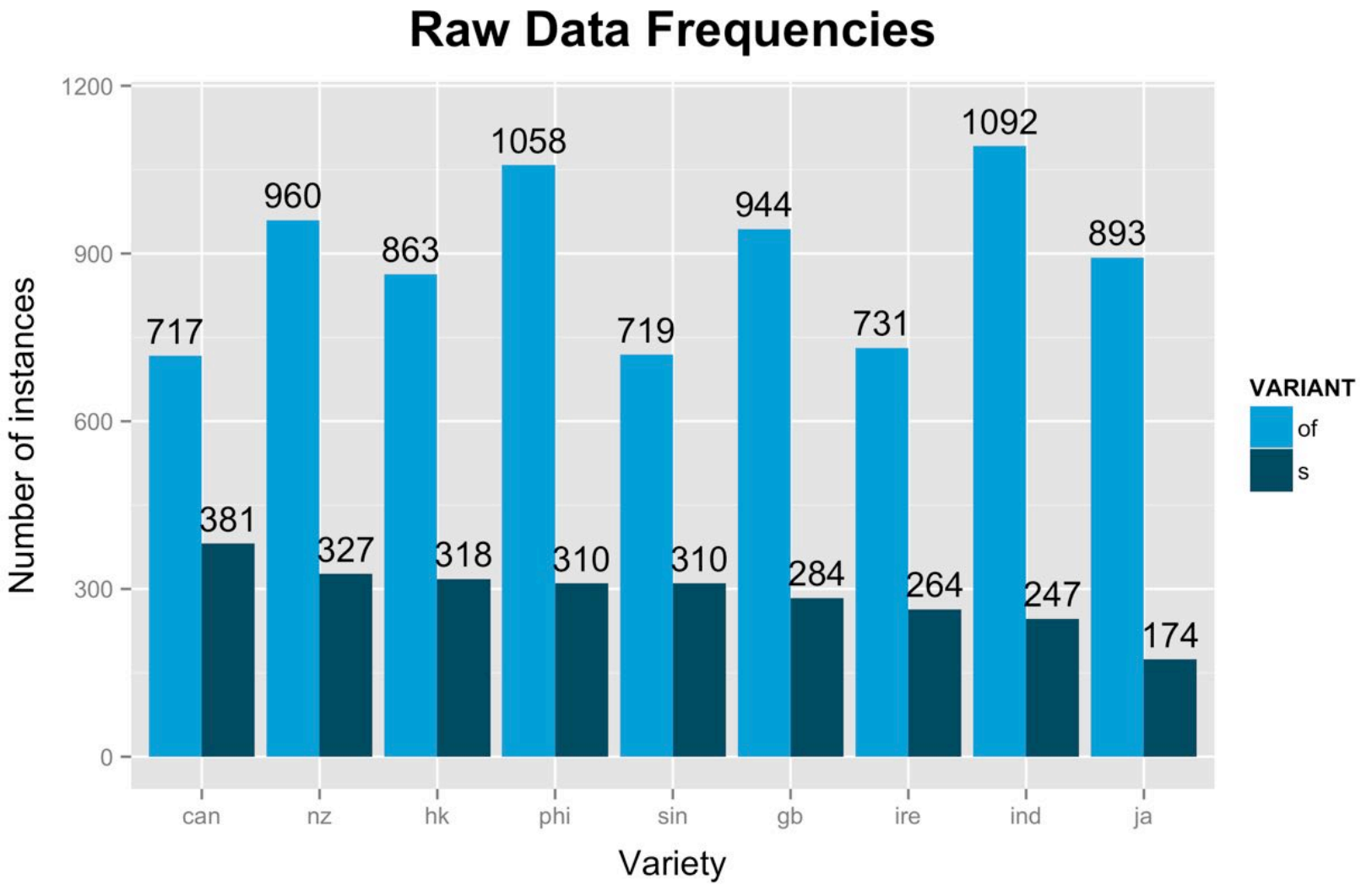
- Semantic relation, e.g. *the girl's mother* vs. *the girl's future* (Rosenbach 2014: 232)
- Definiteness, e.g. *Mary's body* vs. *the woman's body* (Rosenbach 2014: 232)
- Rhythm
- Persistence (i.e. priming)

DATA

CORPORA: ICE-9

- Sample of more than 10,000 interchangeable genitives

RAW FREQUENCIES



REGRESSION MODEL

MIXED-EFFECTS MODELS

- Random effects (REs)
  - Predictors that are special to the sample (but cf. Gelman and Hill 2007: 245–246)
  - It is crucial to account for idiosyncrasies of speakers and corpus structure (Gries 2015)
  - Varying intercepts for speakers nested into GENRE FINE (*ScriptedMono*, *PrivatDia*, etc.) nested into GENRE COARSE (*monologue*, *dialogue*, etc.)
- Model selection according to guidelines in Zuur et al. (2009: ch. 5) and Gries (2015)
- Bootstrap validation (Baayen 2008: 283)

SIGNIFICANT PREDICTORS

	Chisq	Df	Pr(>Chisq)	
POR_ANIMACY	969.9260	1	<2.2e-16	***
MODE	0.1602	1	0.6252034	
POR_FINAL_SIBILANCY	188.2622	1	<2.2e-16	***
POR_GIVENNESS	3.7867	1	0.0448036	*
POR_LENGTH	682.2516	1	<2.2e-16	***
PUM_LENGTH	347.0055	1	<2.2e-16	***
POR_THEMATICITY	4.5525	1	0.0343716	*
TTR	8.6477	1	0.0025249	**
VARIETY	40.0732	8	1.461e-06	***
POR_ANIMACY:VARIETY	67.2192	8	3.813e-12	***
MODE:VARIETY	15.4784	8	0.0342180	*
POR_FINAL_SIBILANCY:VARIETY	31.4791	8	5.667e-05	***
PUM_LENGTH:VARIETY	28.1270	8	0.0002298	***

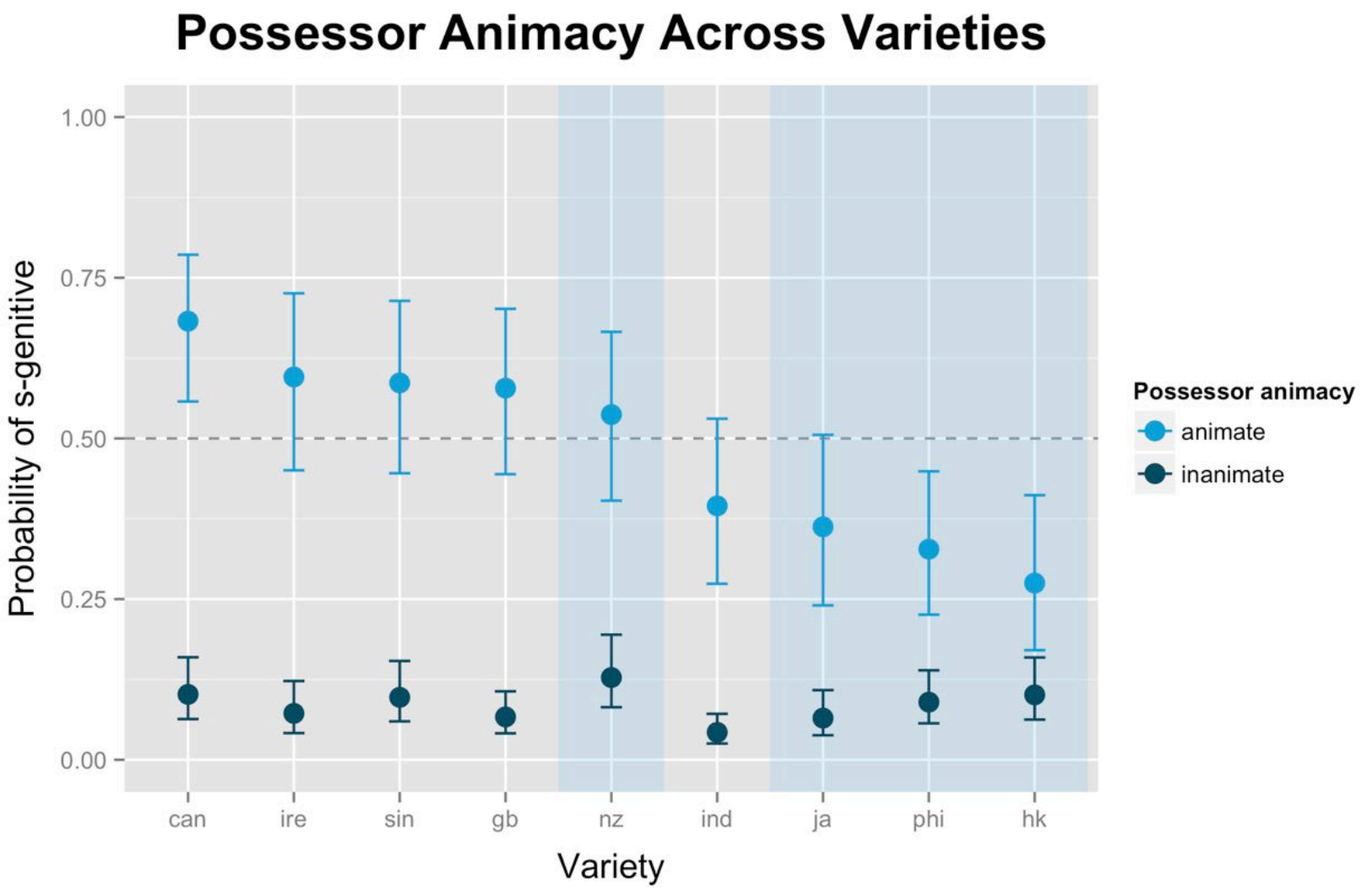
Correct predictions: 87.01%, C-value: 0.7996

FIXED EFFECTS

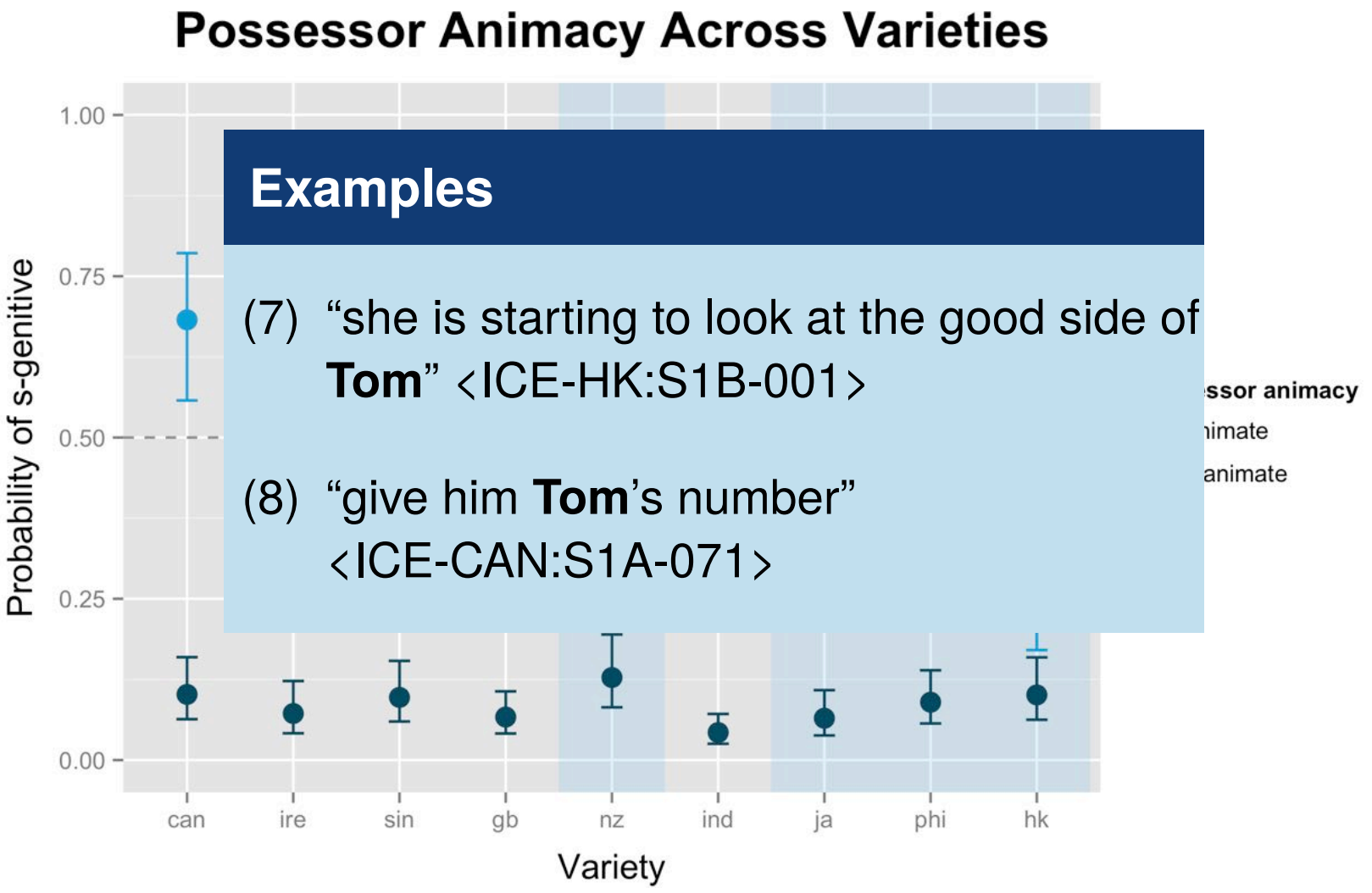
Regression coefficients

Fixed effects	Estimate	Odds	p-value	
POR_ANIMACYa	3.13	23.05	<2e-16	***
POR_FINAL_SIBILANCYtrue	-1.01	0.36	6.43e-05	***
POR_GIVENNESSgiven	0.16	1.17	0.04	*
POR_LENGTH_CHARS_LOG	-1.75	0.17	<2e-16	***
PUM_LENGTH_CHARS_LOG	0.76	2.15	2.65e-05	***
POR_THEMATICITY_LOG	0.07	1.08	0.03	*
TTR	1.54	4.69	0.002	**
VARIETYhk	1.17	3.23	0.0004	***
VARIETYnz	0.68	1.99	0.043	*
VARIETYphi	0.78	2.18	0.018	*
VARIETYsin	0.78	2.19	0.02	*

EFFECTS: ANIMACY * VARIETY



EFFECTS: ANIMACY * VARIETY

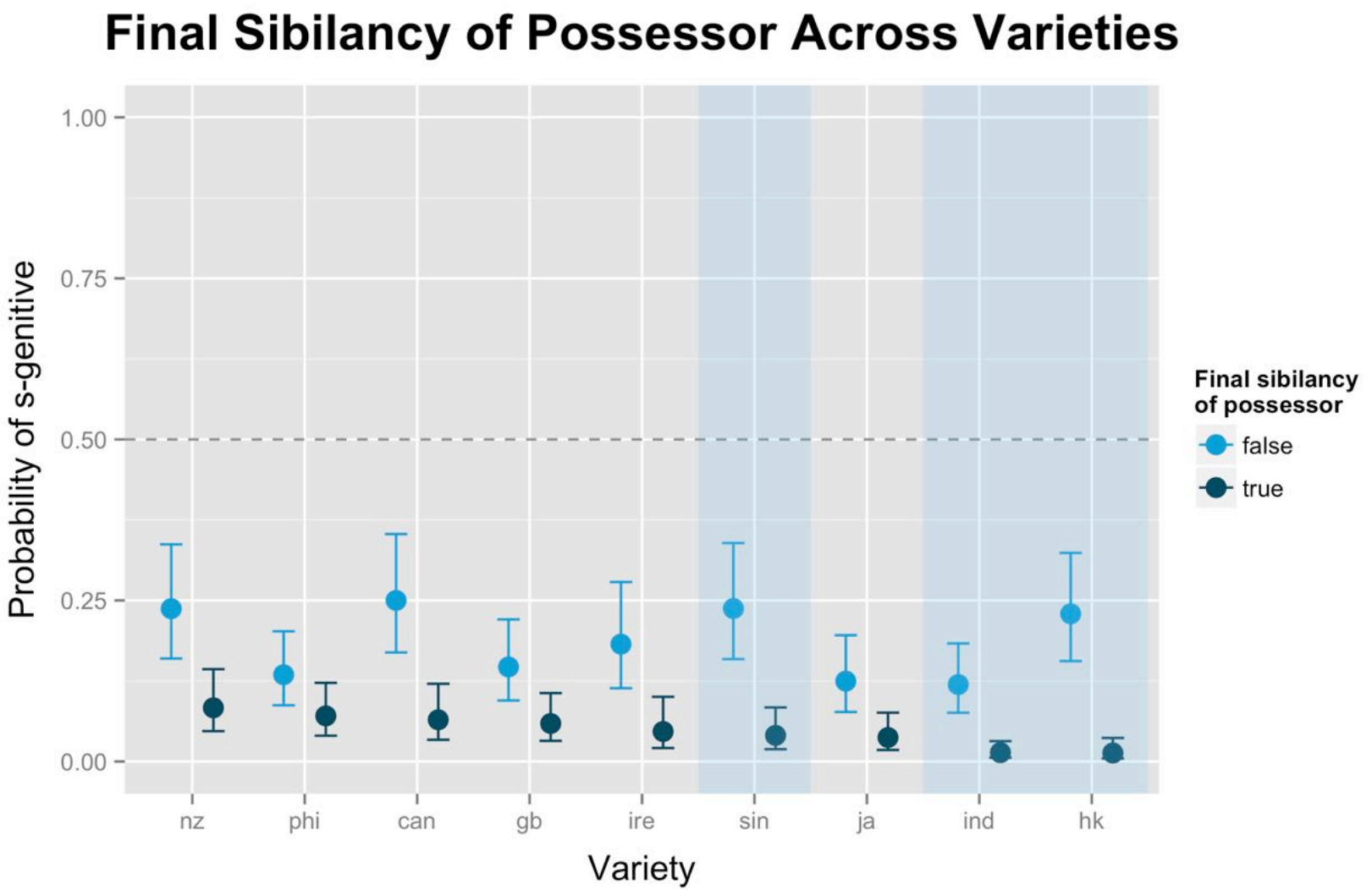


Examples

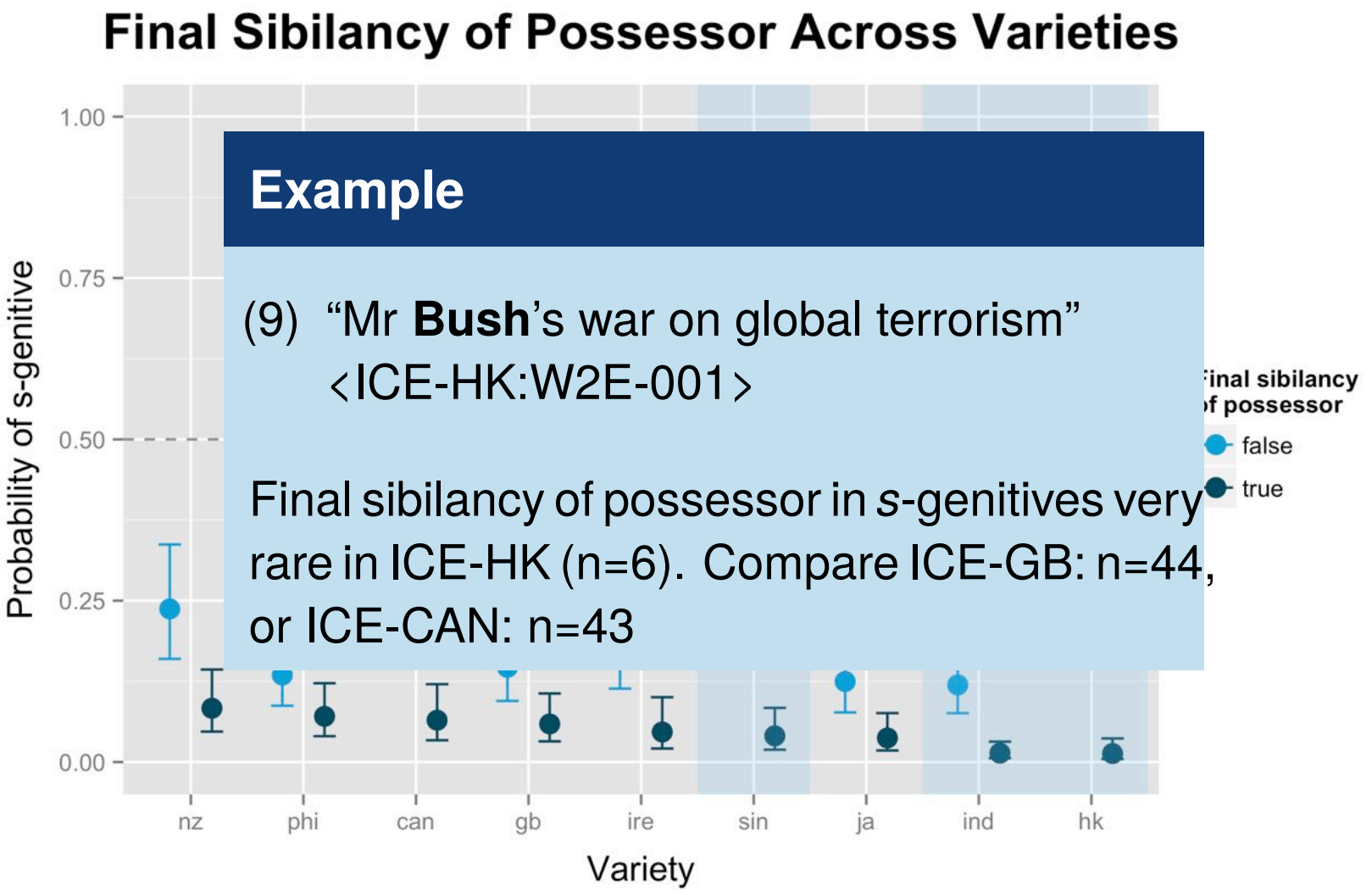
(7) "she is starting to look at the good side of **Tom**" ‹ICE-HK:S1B-001›

(8) "give him **Tom**'s number" ‹ICE-CAN:S1A-071›

EFFECTS: FINAL SIBILANCY * VARIETY



EFFECTS: FINAL SIBILANCY * VARIETY

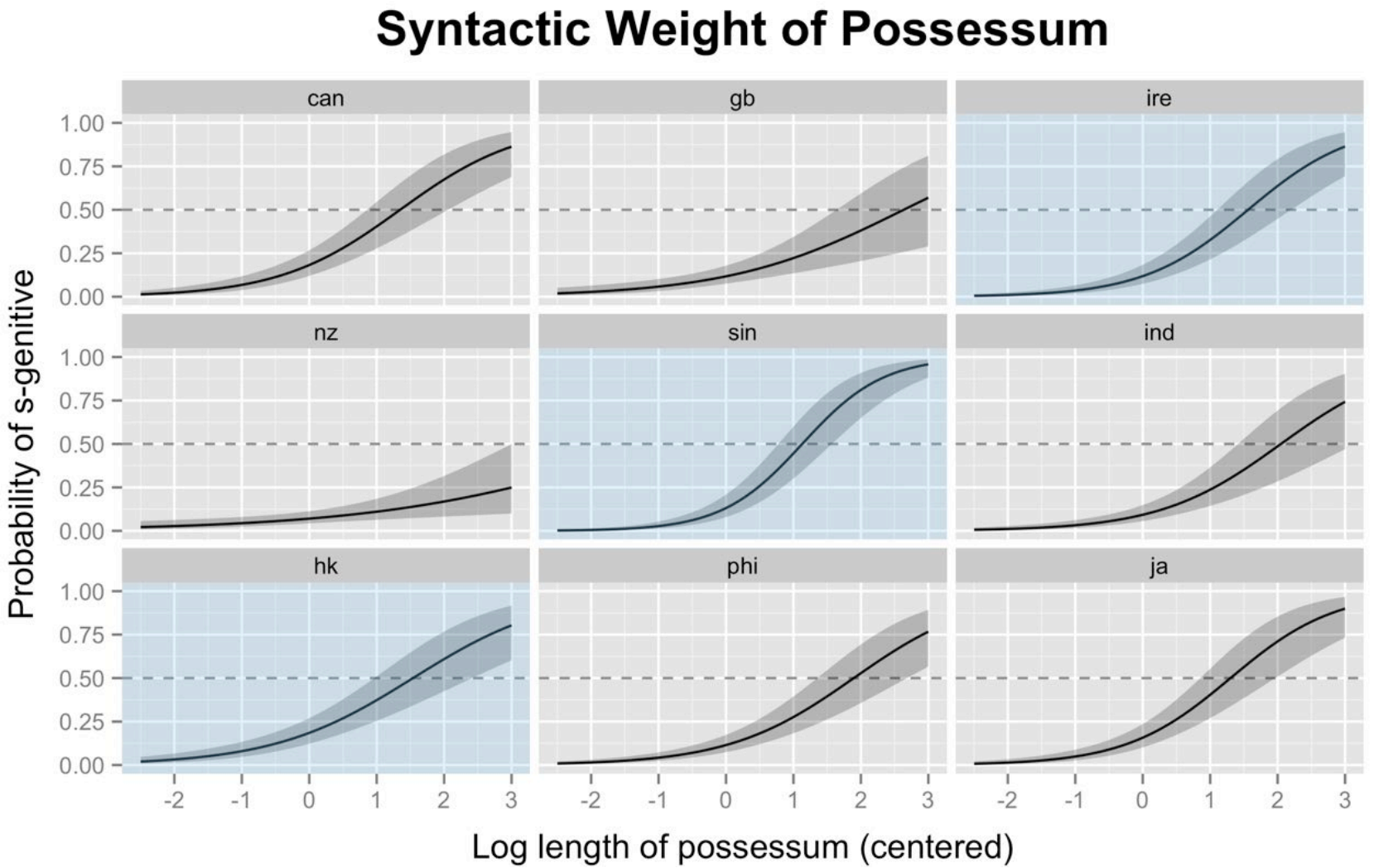


Example

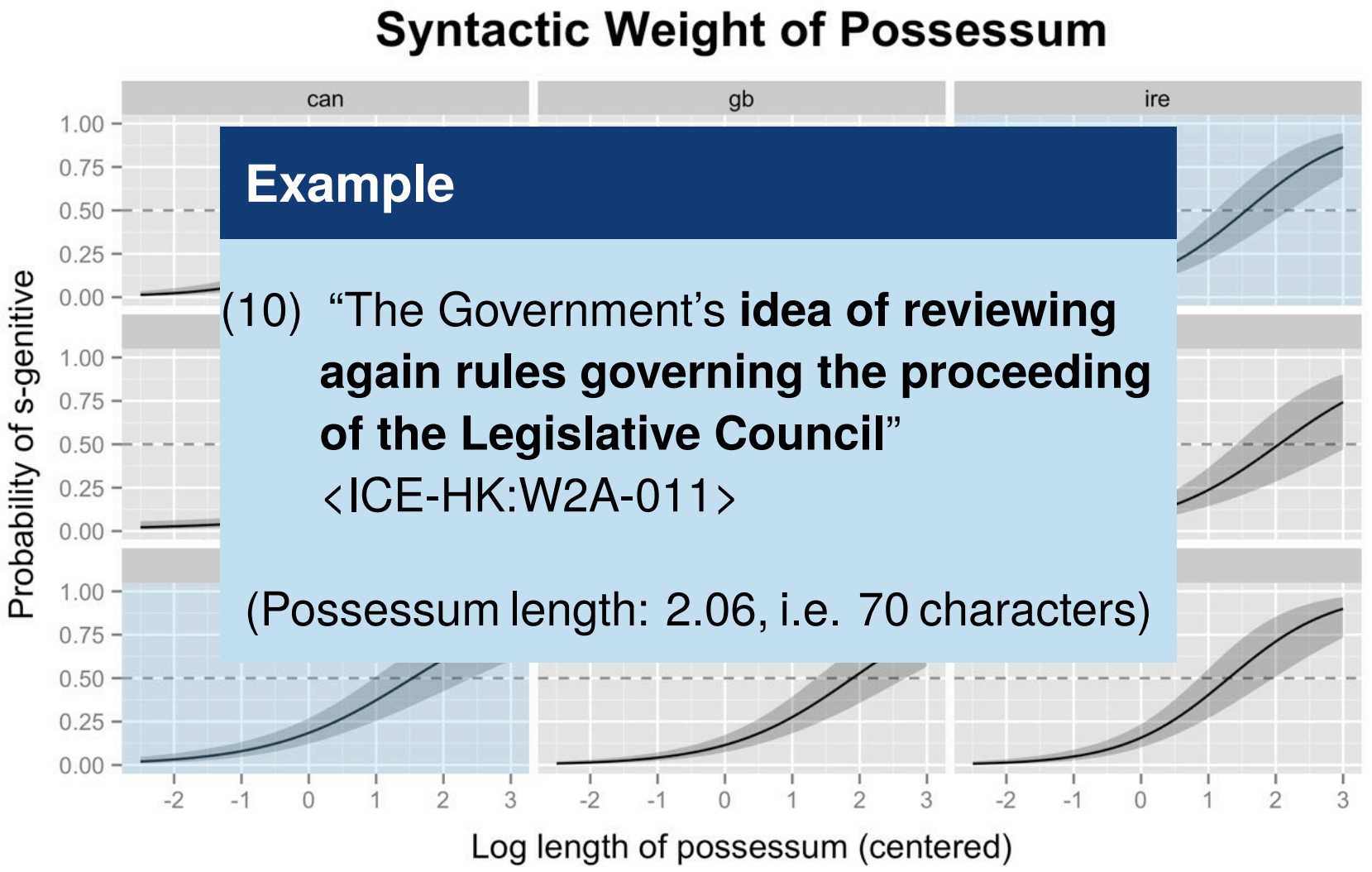
(9) "Mr **Bush**'s war on global terrorism"
‹ICE-HK:W2E-001›

Final sibilancy of possessor in *s*-genitives very rare in ICE-HK (n=6). Compare ICE-GB: n=44, or ICE-CAN: n=43

EFFECTS: POSSESSUM LENGTH * VARIETY



EFFECTS: POSSESSUM LENGTH * VARIETY

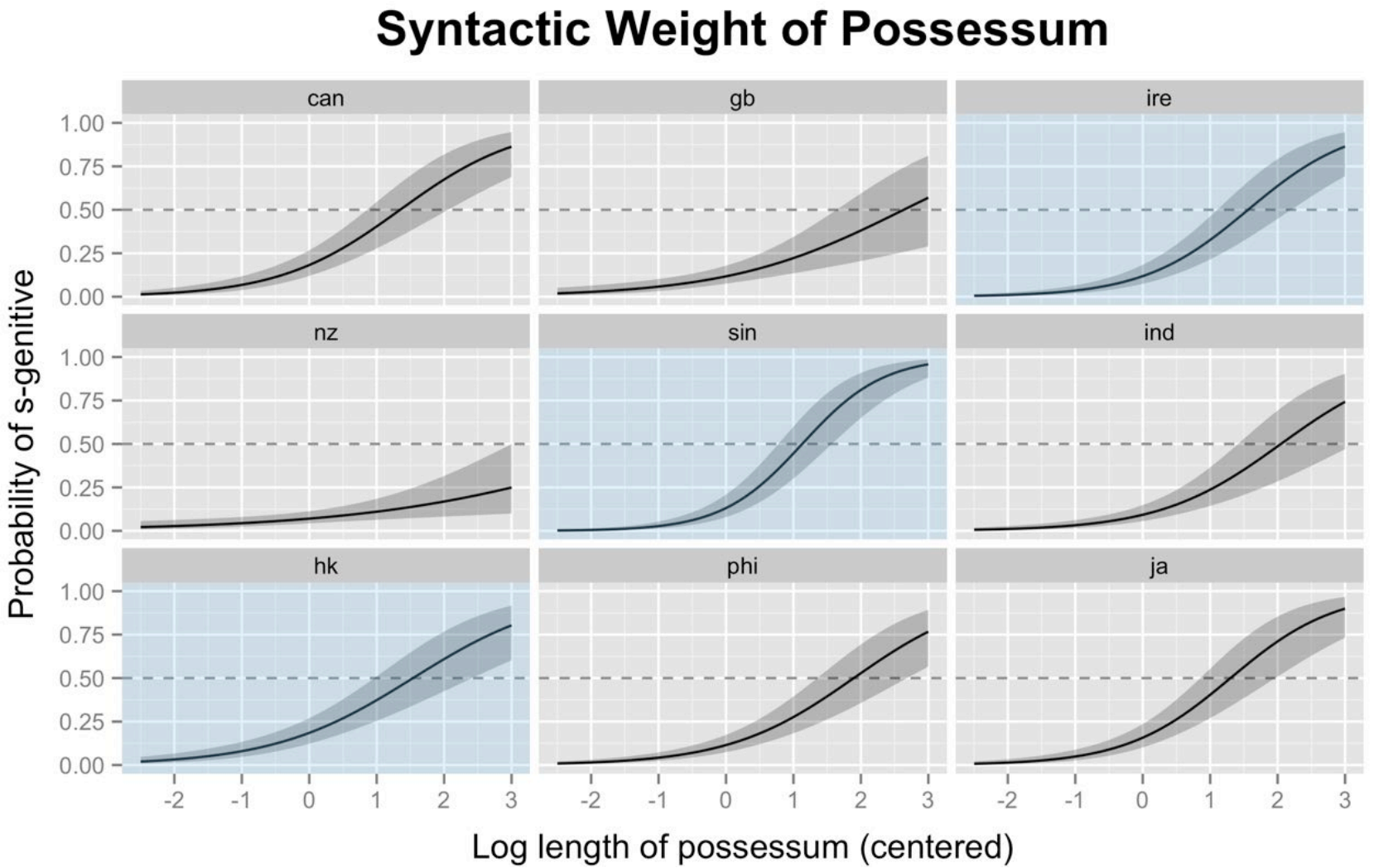


Example

(10) "The Government's **idea of reviewing again rules governing the proceeding of the Legislative Council**" <ICE-HK:W2A-011>

(Possessum length: 2.06, i.e. 70 characters)

EFFECTS: POSSESSUM LENGTH * VARIETY



INTERACTIONS

Regression coefficients

Interactions	Estimate	Odds	p-value	
POR_ANIMACYa:VARhk	-1.81	0.16	3.29e-08	***
POR_ANIMACYa:VARja	-0.89	0.41	0.009	**
POR_ANIMACYa:VARnz	-0.94	0.39	0.002	**
POR_ANIMACYa:VARphi	-1.45	0.23	6.46e-07	***
MODEspoken:VARphi	-1.25	0.28	0.005	**
FINAL_SIBILANCYtrue:VARhk	-2.10	0.12	0.0002	***
FINAL_SIBILANCYtrue:VARind	-1.26	0.28	0.005	**
FINAL_SIBILANCYtrue:VARsin	-0.99	0.37	0.02	*
PUM_LENGTH_LOG:VARhk	0.52	1.68	0.03	*
PUM_LENGTH_LOG:VARire	0.91	2.48	0.0003	***
PUM_LENGTH_LOG:VARsin	0.52	1.69	0.04	*

CONCLUSION

- There are interesting varietal differences in predictors that constrain the genitive alternation
- These constraints point to phonological (final sibilancy), cultural (role of animacy), and cognitive (syntactic weight) differences between the varieties
- Evolutionary status of varieties correlates with strength of constraints

Next steps

- Extract another 5,000 genitives from GloWbE
- More fine-grained animacy classification following Wolk et al. (2013) plus human-animal distinction
- Include other important factors, e.g. semantic relation, definiteness, or persistence

BIBLIOGRAPHY

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statisics Using R*. Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Pearson.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Grafmiller, J. (2014). Variation in english genitives across modality and genres. *English Language and Linguistics 18*(3), 472–496.

Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora 10*(1), 95–125.

Hinrichs, L. and B. Szmrecsanyi (2007). Recent changes in the function and frequency of standard english genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics 11*(3), 437–474.

Murphy, R. (2012). *English Grammar in Use. A self-study reference and practice book for intermetdiate learners of English.* (4 ed.). Cambridge University Press.

Rosenbach, A. (2002). *Genitive Variation in English. Conceptual Factors in Synchronic and Diachronic Studies*. Mouton de Gruyter.

Rosenbach, A. (2014). English genitive variation – the state of the art. *English Language and Linguistics 18*, 215–262.

Schneider, E. (2007). *Postcolonial English: Varieties Around the World*. Cambridge University Press.

Wolk, C., J. Bresnan, A. Rosenbach, and B. Szmrecsanyi (2013). Dative and genitive variability in late modern english. *Diachronica 30*(3), 382–419.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer.

THANK YOU FOR LISTENING

Comments?
Ideas?
Questions?

APPENDIX A: SUBSTRATE LANGUAGES